

2

RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.

Application Serial Number: 10/511,098
Source: PG/10
Date Processed by STIC: 10/21/04

ENTERED



PCT

RAW SEQUENCE LISTING

DATE: 10/21/2004

PATENT APPLICATION: US/10/511,098

TIME: 08:24:32

Input Set : A:\Q83564 Sequence Listing.txt

Output Set: N:\CRF4\10212004\J511098.raw

```

3 <110> APPLICANT: SEKISUI CHEMICAL CO., LTD.
4     MARINE BIOTECHNOLOGY INSTITUTE CO., LTD.
5     IDENO, Akira
6     MARUYAMA, Tadashi
7     FURUTANI, Masahiro
9 <120> TITLE OF INVENTION: EXPRESSION VECTOR, HOST, FUSED PROTEIN, PROCESS FOR
PRODUCING
10     FUSED PROTEIN AND PROCESS FOR PRODUCING PROTEIN
12 <130> FILE REFERENCE: Q83564
C--> 14 <140> CURRENT APPLICATION NUMBER: US/10/511,098
C--> 14 <141> CURRENT FILING DATE: 2004-10-14
14 <150> PRIOR APPLICATION NUMBER: PCT/JP2003/008020
15 <151> PRIOR FILING DATE: 2003-06-25
17 <150> PRIOR APPLICATION NUMBER: JP 2002-185020
18 <151> PRIOR FILING DATE: 2002-06-25
20 <160> NUMBER OF SEQ ID NOS: 30
22 <170> SOFTWARE: PatentIn version 3.3
24 <210> SEQ ID NO: 1
25 <211> LENGTH: 257
26 <212> TYPE: PRT
27 <213> ORGANISM: Pyrococcus horikoshii
29 <400> SEQUENCE: 1
31 Met Lys Val Glu Arg Gly Asp Val Ile Arg Leu His Tyr Thr Gly Arg
32 1           5           10           15
35 Val Lys Glu Thr Gly Gln Ile Phe Asp Thr Thr Tyr Glu Glu Val Ala
36           20           25           30
39 Lys Glu Ala Gly Ile Tyr Asn Pro Lys Gly Ile Tyr Gly Pro Val Pro
40           35           40           45
43 Ile Ile Val Gly Ala Gly His Val Ile Ser Gly Leu Asp Lys Arg Leu
44           50           55           60
47 Val Gly Leu Glu Val Gly Lys Lys Tyr Thr Leu Glu Val Pro Pro Glu
48 65           70           75           80
51 Glu Gly Phe Gly Leu Arg Asp Pro Lys Leu Ile Lys Val Phe Thr Met
52           85           90           95
55 Gly Gln Phe Arg Lys Gln Gly Ile Val Pro Phe Pro Gly Leu Glu Val
56           100          105          110
59 Glu Val Thr Thr Asp Asn Gly Arg Lys Met Lys Gly Arg Val Ile Thr
60           115          120          125
63 Val Ser Gly Gly Arg Val Arg Val Asp Phe Asn His Pro Leu Ala Gly
64           130          135          140
67 Lys Thr Leu Ile Tyr Glu Val Glu Ile Val Glu Lys Ile Glu Asp Pro
68 145          150          155          160
71 Ile Glu Lys Ile Lys Ala Leu Ile Glu Leu Arg Leu Pro Met Ile Asp
72           165          170          175

```

RAW SEQUENCE LISTING

DATE: 10/21/2004

PATENT APPLICATION: US/10/511,098

TIME: 08:24:32

Input Set : A:\Q83564 Sequence Listing.txt

Output Set: N:\CRF4\10212004\J511098.raw

```

75 Arg Asp Lys Val Ile Ile Glu Val Gly Glu Lys Asp Val Lys Val Asn
76          180          185          190
79 Phe Gly Glu Gln Asp Val Asp Pro Lys Thr Leu Ile Leu Gly Glu Ile
80          195          200          205
83 Leu Leu Glu Ser Asp Ile Lys Phe Leu Gly Tyr Glu Lys Val Glu Phe
84          210          215          220
87 Lys Pro Ser Val Glu Glu Leu Leu Arg Pro Lys Gln Glu Glu Pro Val
88 225          230          235          240
91 Glu Glu Glu Lys Lys Glu Glu Gln Glu Glu Ser Glu Glu Ala Gln Ser
92          245          250          255
95 Ser
99 <210> SEQ ID NO: 2
100 <211> LENGTH: 157
101 <212> TYPE: PRT
102 <213> ORGANISM: Methanococcus jannaschii
104 <400> SEQUENCE: 2
106 Leu Ile Asn Leu Ile Lys Lys Gly Asp Tyr Val Lys Val Asp Tyr Ile
107 1          5          10          15
110 Leu Glu Val Asp Gly Lys Val Ile Asp Thr Ser Ile Glu Glu Val Ala
111          20          25          30
114 Lys Glu Asn Lys Ile Tyr Tyr Pro Glu Arg Glu Tyr Glu Pro Ile Gly
115          35          40          45
118 Phe Ile Val Gly Asn Gly Glu Leu Ile Glu Gly Phe Glu Glu Ala Val
119          50          55          60
122 Ile Gly Met Glu Val Gly Glu Glu Lys Thr Val Thr Ile Pro Pro Glu
123 65          70          75          80
126 Lys Gly Tyr Gly Leu Arg Asp Glu Arg Leu Ile Gln Glu Ile Pro Lys
127          85          90          95
130 Glu Met Phe Ala Asp Ala Asp Phe Glu Pro Gln Glu Gly Met Leu Ile
131          100          105          110
134 Leu Ala Ser Gly Ile Pro Ala Lys Ile Ile Lys Val Thr Asp Asp Thr
135          115          120          125
138 Val Thr Leu Asp Phe Asn His Glu Leu Ala Gly Lys Glu Leu Lys Phe
139          130          135          140
142 Thr Ile Lys Val Arg Asp Val Gln Pro Ala Glu Ser Glu
143 145          150          155
146 <210> SEQ ID NO: 3
147 <211> LENGTH: 432
148 <212> TYPE: PRT
149 <213> ORGANISM: Escherichia coli
151 <400> SEQUENCE: 3
153 Met Gln Val Ser Val Glu Thr Thr Gln Gly Leu Gly Arg Arg Val Thr
154 1          5          10          15
157 Ile Thr Ile Ala Ala Asp Ser Ile Glu Thr Ala Val Lys Ser Glu Leu
158          20          25          30
161 Val Asn Val Ala Lys Lys Val Arg Ile Asp Gly Phe Arg Lys Gly Lys
162          35          40          45
165 Val Pro Met Asn Ile Val Ala Gln Arg Tyr Gly Ala Ser Val Arg Gln
166          50          55          60

```

RAW SEQUENCE LISTING

DATE: 10/21/2004

PATENT APPLICATION: US/10/511,098

TIME: 08:24:32

Input Set : A:\Q83564 Sequence Listing.txt

Output Set: N:\CRF4\10212004\J511098.raw

```

169 Asp Val Leu Gly Asp Leu Met Ser Arg Asn Phe Ile Asp Ala Ile Ile
170 65 70 75 80
173 Lys Glu Lys Ile Asn Pro Ala Gly Ala Pro Thr Tyr Val Pro Gly Glu
174 85 90 95
177 Tyr Lys Leu Gly Glu Asp Phe Thr Tyr Ser Val Glu Phe Glu Val Tyr
178 100 105 110
181 Pro Glu Val Glu Leu Gln Gly Leu Glu Ala Ile Glu Val Glu Lys Pro
182 115 120 125
185 Ile Val Glu Val Thr Asp Ala Asp Val Asp Gly Met Leu Asp Thr Leu
186 130 135 140
189 Arg Lys Gln Gln Ala Thr Trp Lys Glu Lys Asp Gly Ala Val Glu Ala
190 145 150 155 160
193 Glu Asp Arg Val Thr Ile Asp Phe Thr Gly Ser Val Asp Gly Glu Glu
194 165 170 175
197 Phe Glu Gly Gly Lys Ala Ser Asp Phe Val Leu Ala Met Gly Gln Gly
198 180 185 190
201 Arg Met Ile Pro Gly Phe Glu Asp Gly Ile Lys Gly His Lys Ala Gly
202 195 200 205
205 Glu Glu Phe Thr Ile Asp Val Thr Phe Pro Glu Glu Tyr His Ala Glu
206 210 215 220
209 Asn Leu Lys Gly Lys Ala Ala Lys Phe Ala Ile Asn Leu Lys Lys Val
210 225 230 235 240
213 Glu Glu Arg Glu Leu Pro Glu Leu Thr Ala Glu Phe Ile Lys Arg Phe
214 245 250 255
217 Gly Val Glu Asp Gly Ser Val Glu Gly Leu Arg Ala Glu Val Arg Lys
218 260 265 270
221 Asn Met Glu Arg Glu Leu Lys Ser Ala Ile Arg Asn Arg Val Lys Ser
222 275 280 285
225 Gln Ala Ile Glu Gly Leu Val Lys Ala Asn Asp Ile Asp Val Pro Ala
226 290 295 300
229 Ala Leu Ile Asp Ser Glu Ile Asp Val Leu Arg Arg Gln Ala Ala Gln
230 305 310 315 320
233 Arg Phe Gly Gly Asn Glu Lys Gln Ala Leu Glu Leu Pro Arg Glu Leu
234 325 330 335
237 Phe Glu Glu Gln Ala Lys Arg Arg Val Val Val Gly Leu Leu Leu Gly
238 340 345 350
241 Glu Val Ile Arg Thr Asn Glu Leu Lys Ala Asp Glu Glu Arg Val Lys
242 355 360 365
245 Gly Leu Ile Glu Glu Met Ala Ser Ala Tyr Glu Asp Pro Lys Glu Val
246 370 375 380
249 Ile Glu Phe Tyr Ser Lys Asn Lys Glu Leu Met Asp Asn Met Arg Asn
250 385 390 395 400
253 Val Ala Leu Glu Glu Gln Ala Val Glu Ala Val Leu Ala Lys Ala Lys
254 405 410 415
257 Val Thr Glu Lys Glu Thr Thr Phe Asn Glu Leu Met Asn Gln Gln Ala
258 420 425 430
261 <210> SEQ ID NO: 4
262 <211> LENGTH: 1299
263 <212> TYPE: DNA

```

RAW SEQUENCE LISTING

DATE: 10/21/2004

PATENT APPLICATION: US/10/511,098

TIME: 08:24:32

Input Set : A:\Q83564 Sequence Listing.txt

Output Set: N:\CRF4\10212004\J511098.raw

```

264 <213> ORGANISM: Escherichia coli
266 <400> SEQUENCE: 4
267 atgcaagttt cagttgaaac cactcaaggc cttggccgcc gtgtaacgat tactatcgct      60
269 gctgacagca tcgagaccgc tgttaaaagc gagctgggtca acgttgcgaa aaaagtacgt      120
271 attgacggct tccgcaaagg caaagtgcc atgaatatcg ttgctcagcg ttatggcgcg      180
273 tctgtacgcc aggacgttct gggtgacctg atgagccgta acttcattga cgccatcatt      240
275 aaagaaaaaa tcaatccggc tggcgccacc acttatgttc cgggcgaata caagctgggt      300
277 gaagacttca cttactctgt agagtttgaa gtttatccgg aagttgaact gcagggctctg      360
279 gaagcgatcg aagttgaaaa accgatcggt gaagtgaccg acgctgacgt tgacggcatg      420
281 ctggatactc tgcgtaaaca gcaggcgacc tggaaagaaa aagacggcgc tgttgaagca      480
283 gaagaccgcg taaccatcga cttaccgggt tctgtagacg gcgaagagtt cgaaggcggg      540
285 aaagcgtctg atttcgtact ggcgatgggc cagggtcgta tgatcccggg ctttgaagac      600
287 ggtatcaaag gccacaaagc tggcgaagag ttcaccatcg acgtgacctt cccggaagaa      660
289 taccacgcag aaaacctgaa aggtaaagca gcgaaattcg ctatcaacct gaagaaagtt      720
291 gaagagcgtg aactgccgga actgactgca gaattcatca aacgtttcgg cggtgaagat      780
293 ggttccgtag aagggtctgcg cgctgaagtg cgtaaaaaaca tggagcgcgga gctgaagagc      840
295 gccatccgta accgcgttaa gtctcaggcg atcgaagggtc tggtaaaagc taacgacatc      900
297 gacgtaccgg ctgcgctgat cgacagcgaa atcgacgttc tgcgtcgcca ggctgcacag      960
299 cgtttcggtg gcaacgaaaa acaagctctg gaactgccgc gcgaactgtt cgaagaacag      1020
301 gctaaacgcc gcgtagtgtg tggcctgctg ctgggcgaag ttatccgcac caacgagctg      1080
303 aaagctgacg aagagcgcggt gaaaggcctg atcgaagaga tggcttctgc gtacgaagat      1140
305 ccgaaagaag ttatcgagtt ctacagcaaa aacaaagaac tgatggacaa catgcgcaat      1200
307 gttgctctgg aagaacaggc tgttgaagct gtactggcga aagcgaaagt gactgaaaaa      1260
309 gaaaccactt tcaacgagct gatgaaccag caggcgtaa      1299
312 <210> SEQ ID NO: 5
313 <211> LENGTH: 270
314 <212> TYPE: PRT
315 <213> ORGANISM: Escherichia coli
317 <400> SEQUENCE: 5
319 Met Lys Ser Leu Phe Lys Val Thr Leu Leu Ala Thr Thr Met Ala Val
320 1 5 10 15
323 Ala Leu His Ala Pro Ile Thr Phe Ala Ala Glu Ala Ala Lys Pro Ala
324 20 25 30
327 Thr Ala Ala Asp Ser Lys Ala Ala Phe Lys Asn Asp Asp Gln Lys Ser
328 35 40 45
331 Ala Tyr Ala Leu Gly Ala Ser Leu Gly Arg Tyr Met Glu Asn Ser Leu
332 50 55 60
335 Lys Glu Gln Glu Lys Leu Gly Ile Lys Leu Asp Lys Asp Gln Leu Ile
336 65 70 75 80
339 Ala Gly Val Gln Asp Ala Phe Ala Asp Lys Ser Lys Leu Ser Asp Gln
340 85 90 95
343 Glu Ile Glu Gln Thr Leu Gln Ala Phe Glu Ala Arg Val Lys Ser Ser
344 100 105 110
347 Ala Gln Ala Lys Met Glu Lys Asp Ala Ala Asp Asn Glu Ala Lys Gly
348 115 120 125
351 Lys Glu Tyr Arg Glu Lys Phe Ala Lys Glu Lys Gly Val Lys Thr Ser
352 130 135 140
355 Ser Thr Gly Leu Val Tyr Gln Val Val Glu Ala Gly Lys Gly Glu Ala
356 145 150 155 160

```

RAW SEQUENCE LISTING

DATE: 10/21/2004

PATENT APPLICATION: US/10/511,098

TIME: 08:24:32

Input Set : A:\Q83564 Sequence Listing.txt

Output Set: N:\CRF4\10212004\J511098.raw

```

359 Pro Lys Asp Ser Asp Thr Val Val Val Asn Tyr Lys Gly Thr Leu Ile
360                               165                               170                               175
363 Asp Gly Lys Glu Phe Asp Asn Ser Tyr Thr Arg Gly Glu Pro Leu Ser
364                               180                               185                               190
367 Phe Arg Leu Asp Gly Val Ile Pro Gly Trp Thr Glu Gly Leu Lys Asn
368                               195                               200                               205
371 Ile Lys Lys Gly Gly Lys Ile Lys Leu Val Ile Pro Pro Glu Leu Ala
372                               210                               215                               220
375 Tyr Gly Lys Ala Gly Val Pro Gly Ile Pro Pro Asn Ser Thr Leu Val
376 225                               230                               235                               240
379 Phe Asp Val Glu Leu Leu Asp Val Lys Pro Ala Pro Lys Ala Asp Ala
380                               245                               250                               255
383 Lys Pro Glu Ala Asp Ala Lys Ala Ala Asp Ser Ala Lys Lys
384                               260                               265                               270

```

387 <210> SEQ ID NO: 6

388 <211> LENGTH: 813

389 <212> TYPE: DNA

390 <213> ORGANISM: Escherichia coli

392 <400> SEQUENCE: 6

```

393 atgaaatcac tgtttaaagt aacgctgctg gcgaccacaa tggccgttgc cctgcatgca      60
395 ccaatcactt ttgctgctga agctgcaaaa cctgctacaa ctgctgacag caaagcagcg      120
397 ttcaaaaatg acgatcagaa atcagcttat gcactgggtg cttegctggg tcgttacatg      180
399 gaaaactctc ttaaagaaca agaaaaactg ggcatacaaac tggataaaga tcagctgac      240
401 gctggtgttc aggatgcatt tgctgataag agcaaacttt ccgaccaaga gatcgaacag      300
403 actctgcaag cattcgaagc tcgctggaag tcttctgctc aggcgaagat ggaaaaagac      360
405 gcggtctgata acgaagcaaa aggtaaagag taccgcgaga aatttgccaa agagaaagg      420
407 gtgaaaacct cttccactgg tctggtttat caggtagtag aagccggtaa aggcgaagcc      480
409 ccgaaagaca gcgatactgt tgtagtgaac tacaaaggta cgctgatcga cggtaaagag      540
411 ttcgacaact cttacaccgg tgggtgaaccg ctctctttcc gtctggacgg tgttatcccg      600
413 gggtggacag aagggtctgaa gaacatcaag aaaggcggta agatcaaact ggttattcca      660
415 ccagaactgg cttacggcaa agcgggtggt ccggggatcc caccgaattc taccctggtg      720
417 tttgacgtag agctgctgga tgtgaaacca ggcgccgaagg ctgatgcaaa gccggaagct      780
419 gatgcgaaag ccgcagactc tgctaaaaaa taa                                     813

```

422 <210> SEQ ID NO: 7

423 <211> LENGTH: 428

424 <212> TYPE: PRT

425 <213> ORGANISM: Escherichia coli

427 <400> SEQUENCE: 7

```

429 Met Lys Asn Trp Lys Thr Leu Leu Leu Gly Ile Ala Met Ile Ala Asn
430 1                               5                               10                               15
433 Thr Ser Phe Ala Ala Pro Gln Val Val Asp Lys Val Ala Ala Val Val
434                               20                               25                               30
437 Asn Asn Gly Val Val Leu Glu Ser Asp Val Asp Gly Leu Met Gln Ser
438                               35                               40                               45
441 Val Lys Leu Asn Ala Ala Gln Ala Arg Gln Gln Leu Pro Asp Asp Ala
442                               50                               55                               60
445 Thr Leu Arg His Gln Ile Met Glu Arg Leu Ile Met Asp Gln Ile Ile
446 65                               70                               75                               80
449 Leu Gln Met Gly Gln Lys Met Gly Val Lys Ile Ser Asp Glu Gln Leu

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/511,098

DATE: 10/21/2004

TIME: 08:24:33

Input Set : A:\Q83564 Sequence Listing.txt

Output Set: N:\CRF4\10212004\J511098.raw

L:14 M:270 C: Current Application Number differs, Replaced Current Application No

L:14 M:271 C: Current Filing Date differs, Replaced Current Filing Date